1 Monte Carlo Simulation Forecasting

1.1 Overview of Monte Carlo Method

1.1.1 Basic Idea

The basic idea of the Monte Carlo method is to use the "frequency" of an event occurring as an approximate value of the "probability" of that event, under the assurance of the law of large numbers. By designing a random experiment in which the probability of an event is related to some unknown variable, and then repeating the experiment to represent the probability with an approximate frequency value, one can obtain an approximate value for that unknown variable. This method requires a considerable number of trials to achieve an approximate solution, and the more trials conducted, the better the approximation.

1.1.2 Main Steps

The Monte Carlo method consists of three main steps:

1. Describe or Construct the Probability Process: For deterministic problems that do not have random characteristics, it is necessary to artificially construct a probability process.

2. Sampling from Probability Distributions: Generate random variables with known probability distributions using a computer, such as uniform distribution, normal distribution, exponential distribution, Poisson distribution, etc.

3. Establish Various Estimators: After constructing the random probability model and sampling from it, a random variable must be identified as the solution to the problem at hand. Generally, the arithmetic mean of the results from the subsequent random sampling is used as an approximate value of the solution.

1.2 Basic Assumptions and Parameter Estimation

1.2.1 Basic Assumptions

In the simulation, we assume that stock prices follow a geometric Brownian motion. Geometric Brownian motion is a variant of Brownian motion that describes the random movement of an asset's price, where the logarithm of the price follows Brownian motion. This model assumes that the volatility of asset prices is constant and that price changes are continuous.
Assume that the change in stock price $S(t)$ follows the following form of stochastic differential equation:

$$dS(t) = \mu S(t)dt + \sigma S(t)\xi$$

where μ is the expected return on the stock price (i.e., the risk-free growth rate of the stock price). σ represents the volatility of the stock price, ξ is the differential of Brownian motion, which indicates the random change between time t and $t + dt$.

1.2.2 Parameter Estimation

Data Selection We selected the trading data for Ping An Bank (stock code: SZ000001) from January 1, 2018, to April 11, 2024, for simulation analysis. The data is sourced from the Tushare library.

Parameter Calculation

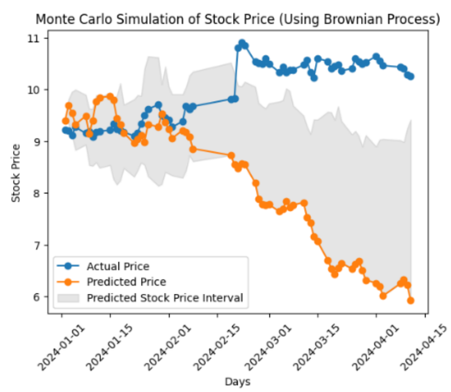
- 5-Year Expected Return(μ): -0.00023683098982947733
- 5-Year Volatility(σ):0.02061630038252004
- 1-Year Expected Return(μ):-0.0015885768444829525
- 1-Year Volatility(σ): 0.013305615290128279

1.3 Simulation Process

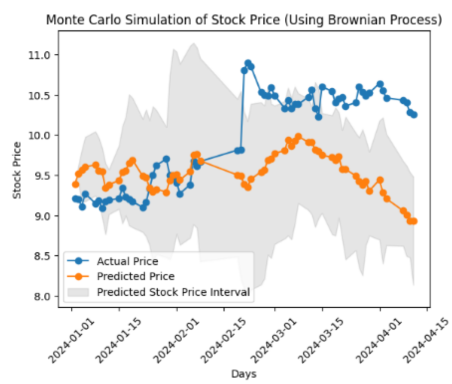
We used the stock price data from January 1, 2018, to December 29, 2023, to predict the stock prices from January 1, 2024, to April 11, 2024.

We employed a method of conducting one simulation and taking the average of five simulations, plotting both the predicted stock price curve and the actual stock price curve.

The prediction results are as follows:



(a) One Simulation



(b) Five Simulation

It can be observed that compared to a single simulation, the prediction results obtained by averaging multiple simulations are more closely aligned with the actual results. The prediction results from a single Monte Carlo simulation exhibit strong randomness, while the method of averaging five simulations helps to offset randomness to a certain extent, thereby improving the accuracy and stability of the predictions. Therefore, in the upcoming simulation studies, we will adopt the method of averaging multiple simulations and also display the range interval to obtain more accurate and comprehensive results.

However, it is noted that there was a significant deviation in the simulations during February. How can this deviation be explained? To this end, the following three hypotheses are proposed:

- The duration of historical data affects the simulation results.
- The number of simulations influences the simulation results.
- Stock price fluctuations are also driven by other (strong) factors.

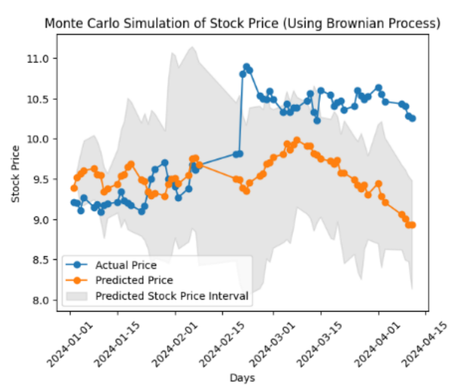
1.4 Analysis and Conclusions

1.4.1 The duration of historical data affects the simulation results, but not much.

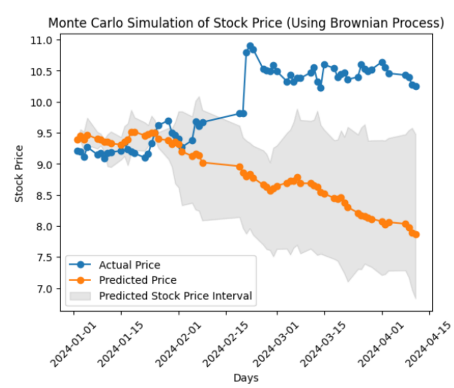
It is hypothesized that the accuracy of predictions may be related to the length of the historical data period selected. Using overly long historical data for simulations may result in an excessive weight on outdated information, while using a very short historical data period may lead to insufficient information, making it difficult to comprehensively capture the characteristics of stock price fluctuations, thereby affecting the accuracy and stability of the predictions.

To validate this hypothesis, we selected five years data (from January 2, 2018, to December 29, 2023) as long-term data, and one year data (from January 3, 2023, to December 29, 2023) as short-term data. We employed the method of averaging five simulations to predict the stock price data from January 2, 2024, to April 11, 2024.

The prediction results are as follows:



(a) Long History Path
Number of Simulation Paths: 5
Historical Interval: 5 years
Simulation Time: Jan 1 2024 to Apr 15 2024



(b) Short History Path
Number of Simulation Paths: 5
Historical Interval: 1 years
Simulation Time: Jan 1 2024 to Apr 15 2024

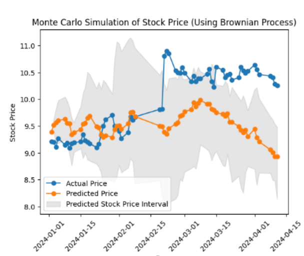
From the simulation results shown in the graph, it can be observed that both the long-term and short-term data predictions exhibit significant deviations from the actual stock prices in February. When the historical period is longer, the predicted price range is more likely to cover the true value of the stock price, but the impact is not significant; therefore, the length of the historical data period is not the main factor affecting the simulation results. In further research, this paper adopts 5 years as the historical period. However, we still need to consider incorporating more influencing factors and making further improvements to the model to enhance its adaptability to sudden events.

1.4.2 The More Simulation We Run, the Better It Converges.

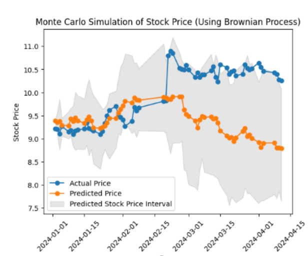
It is hypothesized that the accuracy of predictions is related to the number of simulations conducted. Conducting multiple simulations within a certain range can reduce randomness and improve the model's adaptability to sudden events.

To validate this hypothesis, we selected data from January 2, 2018, to December 29, 2023, and employed both the method of averaging five simulations and the method of averaging ten simulations to predict the stock price data from January 2, 2024, to April 11, 2024.

The prediction results are as follows:



(a) 5 Simulations
Number of Simulation Paths: 5
Historical Interval: 5 years
Simulation Time: Jan 1 2024 to Apr 15 2024



(b) 10 Simulations
Number of Simulation Paths: 10
Historical Interval: 5 years
Simulation Time: Jan 1 2024 to Apr 15 2024

From the simulation results, it can be observed that compared to averaging five simulations, the prediction interval for stock prices when averaging ten simulations generally covers the true value of the stock price. This indicates that appropriately increasing the number of simulations can enhance the stability of the model and reduce the randomness of the prediction results. After experimentation, further increasing the number of predicted stock price paths does not significantly improve prediction accuracy. Considering the cost-benefit balance, this paper will henceforth use the method of averaging ten simulations.

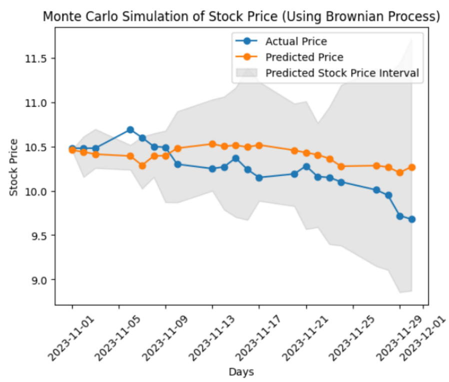
1.4.3 Stock Price Fluctuations Are Driven by Event Factors and Random Walks.

In fact, a closer examination of the time when deviations occurred reveals events such as the liquidity crisis in small-cap stocks in February 2024 and the intervention of state-backed funds to stabilize the market. Such non-market forces are likely the fundamental reasons behind the significant deviation between the simulation predictions and the true stock price in February.

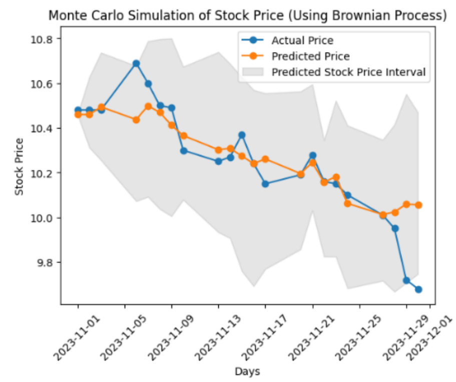
To observe whether the model's prediction accuracy would improve under normal market conditions, we adjusted the prediction period and historical data period, selecting stock price data from January 2, 2018, to October 30, 2023, as long-term data, and stock price data from January 1, 2023, to October 30, 2023, as short-term data. We then predicted the stock data from November 1, 2023, to November 30, 2023, and re-simulated and analyzed during the new prediction period.

The reason for choosing the period from November 1, 2023, to November 30, 2023, for simulation predictions is that the stock market faced many significant events after the start of 2024, and December 2023, as the last month of the year, also experienced some non-market regulatory factors. Therefore, November, which is relatively blank, was selected as the prediction period. The prediction results are astonishingly good, with predicted stock prices on some days nearly overlapping!

Clearly, the impact of unexpected events on stock prices cannot be overlooked. The formation of stock prices is influenced by both event-driven factors and random walk factors. Non-market forces (such as event-driven



(a) If we use Nov 2023 as test period.(1)



(b) If we use Nov 2023 as test period.(2)

factors—company news, market expectations, and macroeconomic changes) can quickly alter market sentiment and investor decision-making. On the other hand, random walk factors reflect the randomness and uncertainty of the market, leading to potentially unpredictable fluctuations in stock prices in the short term.

Therefore, understanding and analyzing these two influencing factors is crucial for the formation of stock prices, helping us better respond to market risks and opportunities. In stock price forecasting and analysis, it is necessary to comprehensively consider these two types of factors and employ appropriate models and methods to accurately capture the patterns of stock price movements.

2 Model Improvement and Expansion

2.1 Introducing the Markov Process

2.1.1 Model Introduction

Monte Carlo simulation, as a static method, can produce corresponding deviations when changes occur over small intervals, and too many partitions can significantly exacerbate these deviations. Therefore, we attempt to introduce Markov Chain Monte Carlo (MCMC) simulation to achieve the goal of dynamic simulation, where the sampling distribution can continuously change as the simulation progresses.

We constructed a Markov chain using the Metropolis-Hastings algorithm to approximate complex probability distributions, setting the transition matrix as $\begin{bmatrix} 0.2 & 0.5 & 0.3 \\ 0.5 & 0.2 & 0.3 \\ 0.5 & 0.3 & 0.2 \end{bmatrix}$. The results obtained after the simulation are shown in the figure below.

To compare the model fitting effect, we set Mean Squared Error (MSE) as the standard for measuring fitting effectiveness, calculating the proximity of the model's predicted values $\hat{Y}$ to the true values Y. The smaller the MSE, the better the fitting effect.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

2.1.2 Fitting Results

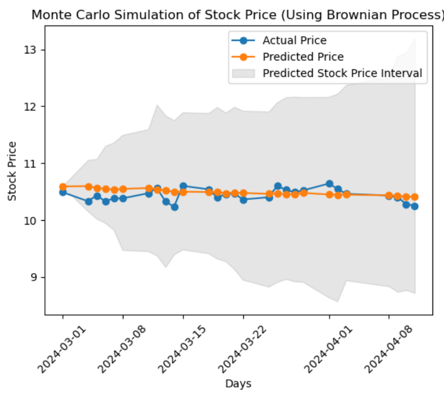
By adjusting the transition matrix, we obtained the following two MCMC simulation results (as shown on page 6). The MSEs were 0.0168 and 0.0159, respectively, while the MSE for the Monte Carlo method simulation was 0.0167, indicating that MCMC did not improve prediction efficiency.

The reasons for this are believed to stem from two factors. Firstly, MCMC has higher requirements for parameter selection, and the parameters we chose differ from the actual situation, necessitating further fine-tuning. Secondly, MCMC converges slowly, and the Markov chain may not have reached a sufficiently converged state. Therefore, we are considering other methods to improve the Monte Carlo simulation.

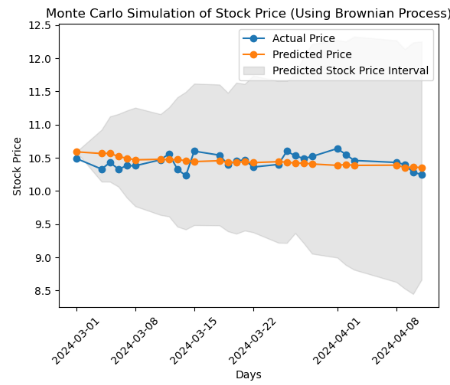
2.2 Introducing the GARCH Model

2.2.1 Model Introduction

The GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model is used to describe the changes in volatility (variance) of time series data. It assumes that the variance at time t is related to the variances and



(a) MCMC Prediction Result 1



(b) MCMC Prediction Result 2

residuals of past periods, representing a form of conditional heteroskedasticity. We adopt a simple GARCH(1,1) model to predict the variance of stock prices.
The expression of the model is as follows:

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

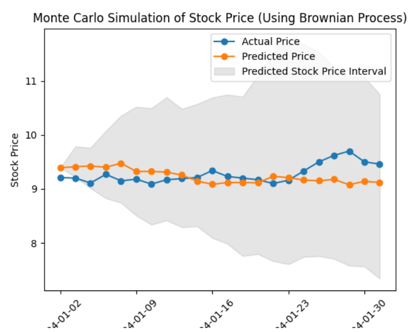
where σ_t^2 is the variance of time t, σ_{t-1}^2 t is the square of the residual (error term) at time $t-1$. ω、α、β are the parameters of the model, representing the constant term, the ARCH effect, and the GARCH effect, respectively.
We multiply the variance predicted by the GARCH model with the random numbers obtained from the Monte Carlo simulation to adjust the model's volatility. That is,

$$\xi' = \xi * \sigma_t$$

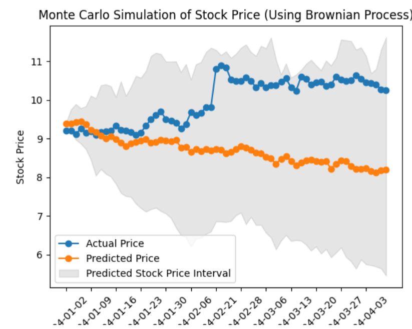
$$dS(t) = \mu S(t)dt + \sigma S(t)\xi'$$

2.2.2 Fitting Results

To explore whether the inclusion of the predicted variance can improve the model's predictive performance under normal market conditions and during sudden events, we used data from January 2, 2018, to December 29, 2023, to predict data for the periods from January 1, 2024, to January 31, 2024, and from January 1, 2024, to April 15, 2024, respectively.



(a) This is a situation where non-market forces are minimal.



(b) This is a situation where there is significant intervention from non-market forces.

3 Conclusions

The prediction results indicate that under normal market conditions, **the model with the inclusion of the predicted variance performs well, and the prediction accuracy has been effectively improved.** Meanwhile, the adjusted model can also mitigate the impact of sudden events on the prediction results to some extent. However, **there remains a significant deviation** between the predicted stock prices and the actual stock prices during sudden events.

This suggests that despite the improvements and adjustments made to the model, using the Monte Carlo method for stock price prediction still faces challenges when confronted with uncertain factors such as unexpected events. **This is also consistent with the underlying logic of the Monte Carlo model—asset prices in financial markets exhibit random walks; therefore, Monte Carlo methods have a strong advantage only when predicting pure market forces.**

Although this method (and its improvements) performs poorly when other forces intervene in the market, its predictive capability regarding fluctuations caused by ”pure market forces” is remarkably high! This sufficiently demonstrates the effectiveness of the method and its profound insights into the financial market.